4/16/2002

name

CSci 124 midterm 2
22 points
There are 10 pages and 10 problems.

1 point 1. Shown below is the ISAM file organization. Show what changes would be caused by inserting a record with key value 403 into this file.

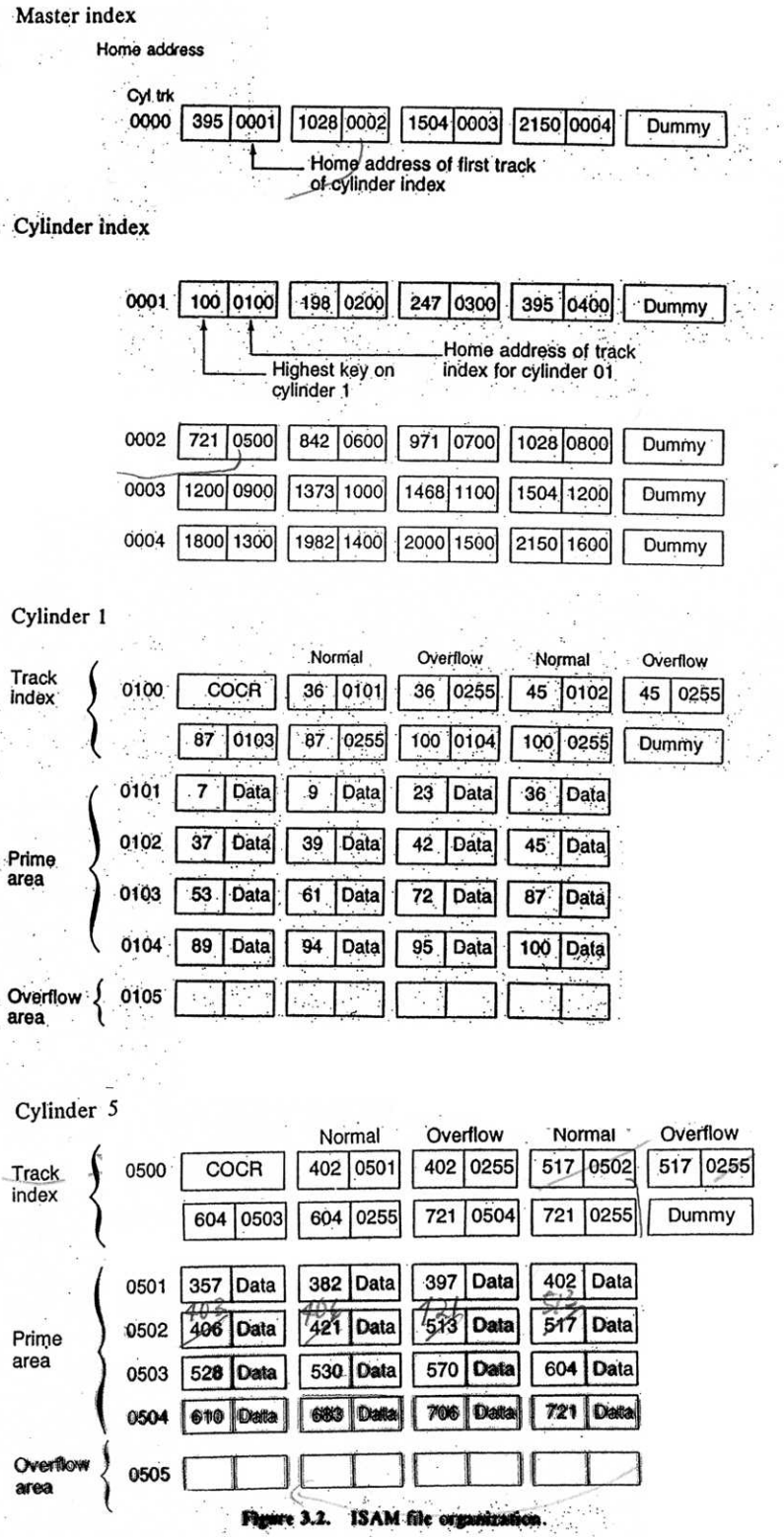

Figure 3.2. ISAM file organization.

2.5 points 2 Keysort

You are given a fixed length record file with a count of the number of records stored in a header record. The file was unsorted. Write the pseudo code for ***keysort*** to sort this file, and reorganize it. Then write the records out in sorted order. Be sure to include the essential steps involving seeking, reading and writing.

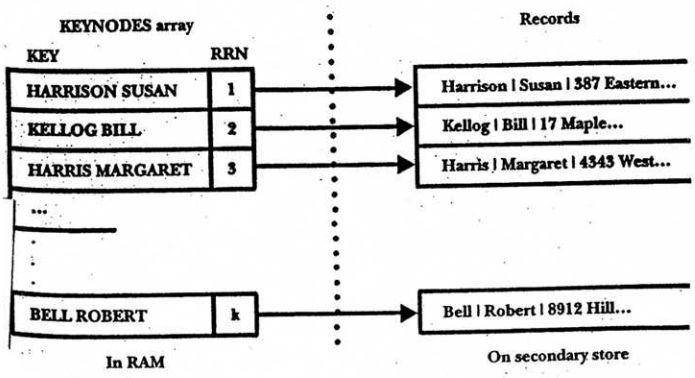

FIGURE 5.14 Conceptual view of KEYNODES array to be used in RAM by internal sort routine, and record array on secondary store.

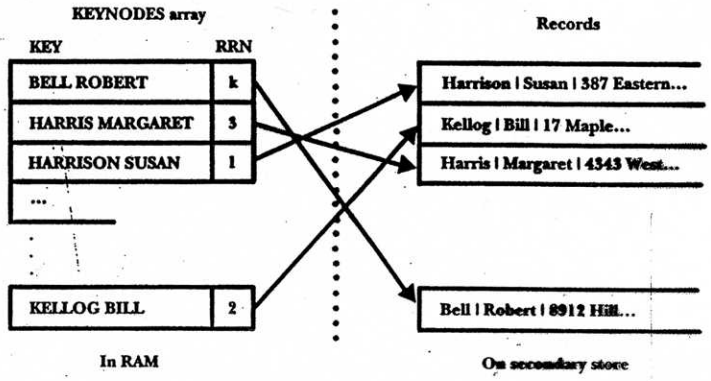

FIGURE 5.15 Conceptual view of KEYNODES array and file after sorting keys in RAM.

1.5 points 3.

(a) Sequential file (1 point)

How many comparisons would be required on average to find a record using sequential search in a 50,000-record disk file?

If the record is not in the file, how many comparisons are required?

If the file is blocked so that 50 records are stored per block, how many disk accesses are required on average?

If only one record is stored per block, how many disk accesses are required on average?

(b) Relative Record Number (0.5 point)

Let n be the RRN of a fixed-length record. Let r be the record size. Give a formula for finding the byte offset of a fixed length record in which the RRN of the first record is 1 rather than 0.

4 points 4. Merge Sort

You are given the following specification of a disk drive:

Average seek time	8 msec
Average rotational delay	3 msec
Maximum transfer rate	6 msec/track or 14500 bytes/msec

In addition, we have the following assumptions:

We have a file with **5,000,000** records, each of which is **100** bytes long. We have **ten** megabyte of memory available as a work area, not counting memory used to hold the program, operating system, I/O buffers, and so forth. We also assume that only one seek and one rotational delay are required for any single sequential access.

We would take this 5,000,000 record file and estimate the time it takes to do a merge sort on the given hypothetical disk drive.

Step 1: Reading Records into memory for sorting and Forming Runs.
Memory is a ten-megabyte input buffer.

Step 2: Writing Sorted Runs out to Disk.

Step 3: Reading Sorted Runs into memory for Merging

Step 4: Writing Sorted File out to Disk.
(We assume that we can allocate **two** 200,000 byte output buffers).

Please fill in the blanks for the following table. The blanks are worth varying points: 1/8, ¼ or ½ points.

	Number of Seeks	Amount Transferred (Megabytes)	Seek + Rotation Time (Seconds)	Transfer Time (Seconds)
Sort: reading				
Sort: writing				
Merge: reading				
Merge: writing				

3 points 6. Polyphase merge sort

1.5 pts (a) Given the recurrence relations at the bottom of the distribution table, express e_n, d_n, c_n, b_n, and a_n in terms of a_i 's.

1.5 pts (b) Complete the distribution table for the 5-way polyphase merge on 6 tapes assumming 65 runs. You may skip the column on 'Final output ..'. Show the polyphase merge for this merge sort. The first two levels are done for you.

Level	T1	T2	T3	T4	T5	Total	Final output will be on
0	1	0	0	0	0	1	T1
1	1	1	1	1	1	5	T6

. . . . . .

n	a_n	b_n	c_n	d_n	e_n	t_n	T(k)
$n+1$	$a_n + b_n$	$a_n + c_n$	$a_n + d_n$	$a_n + e_n$	a_n	$t_n + 4a_n$	$T(k-1)$

3 points 5. Two step merge.

Suppose we have **50,000,000** records instead of 5,000,000 records (with the same specifications as in problem 4). We have 500 runs. Suppose we could merge them as 20 sets of 25 runs.

1st merge step: reading

(i) How many seeks? (1 point)

(ii) What is the seek + rotation time in seconds (total for this part)? (0.5 point)

2nd merge step: reading

(i) How many seeks? (1 point)

(ii) What is the seek + rotation time in seconds(total for this part)? (0.5 point)

2 points 7. Polyphase Merge (2 points, No partial credits for part i)
Given T = 3 tapes, and P = 2, and the initial distribution of runs for the contents of T1 and T2 for phase 1:

(i) Complete the following table showing the contents of T1, T2 and T3 at each phase. (1 ½ points)

(ii) Calculate the total number of "passes" (Knuth's convention). (1/2 point)

	Content of T1	Content of T2	Contents of T3
Phase 1	1,1,1,1,1,1,1,1,1,1,1,1,1,1,1	1,1,1,1,1,1,1,1	

0.5 point 8. Given a multiway tree of $n = 10^6$ nodes, and a constant branching factor of $s = 1000$. What is the average search path length?

2 points 9. Construct a k-d tree with $k = 3$ that results from inserting the following key values in the order from left to right.

(9,F,3) (10,F,4) (8,F,0) (8,B,6) (6,S,3) (12,D,7) (7,B,6) (6,B,14) (10,A,5) (12,F,5)

2.5 points 9. Given fig. 4.20 on page 10 and the definitions of level(i), P, C(P), NLEFT(P, x_i), and X, write a search algorithm for the C-trie.

level(i) - base address of node on level i;
(starting address of leftmost node at level (i)) - 1.

level(1) = 0
level(2) = 1
level(3) = 6
level(4) = 13
level(5) = 14

P - pointer to a node;
U(P), B(P), K(P), C(P), BKC(P);
C(P) - calculate # of branching factor of the nodes to the left;

$X = X_1 \ldots X_k$, $0 \le X_i < m$ for $1 \le i < k$ and
$X_k = 0$ indicates end of key;.
x_i is the bit corresponding to X_i.

NLEFT(P,x_i) - # of 1 bits to the left of and including the x_i bit in K(P) in the node P;
K(P,x_i) - x_ith field of K(P); K(P,0) is the same as B(P);

algorithm search-trie

Root is at level 1

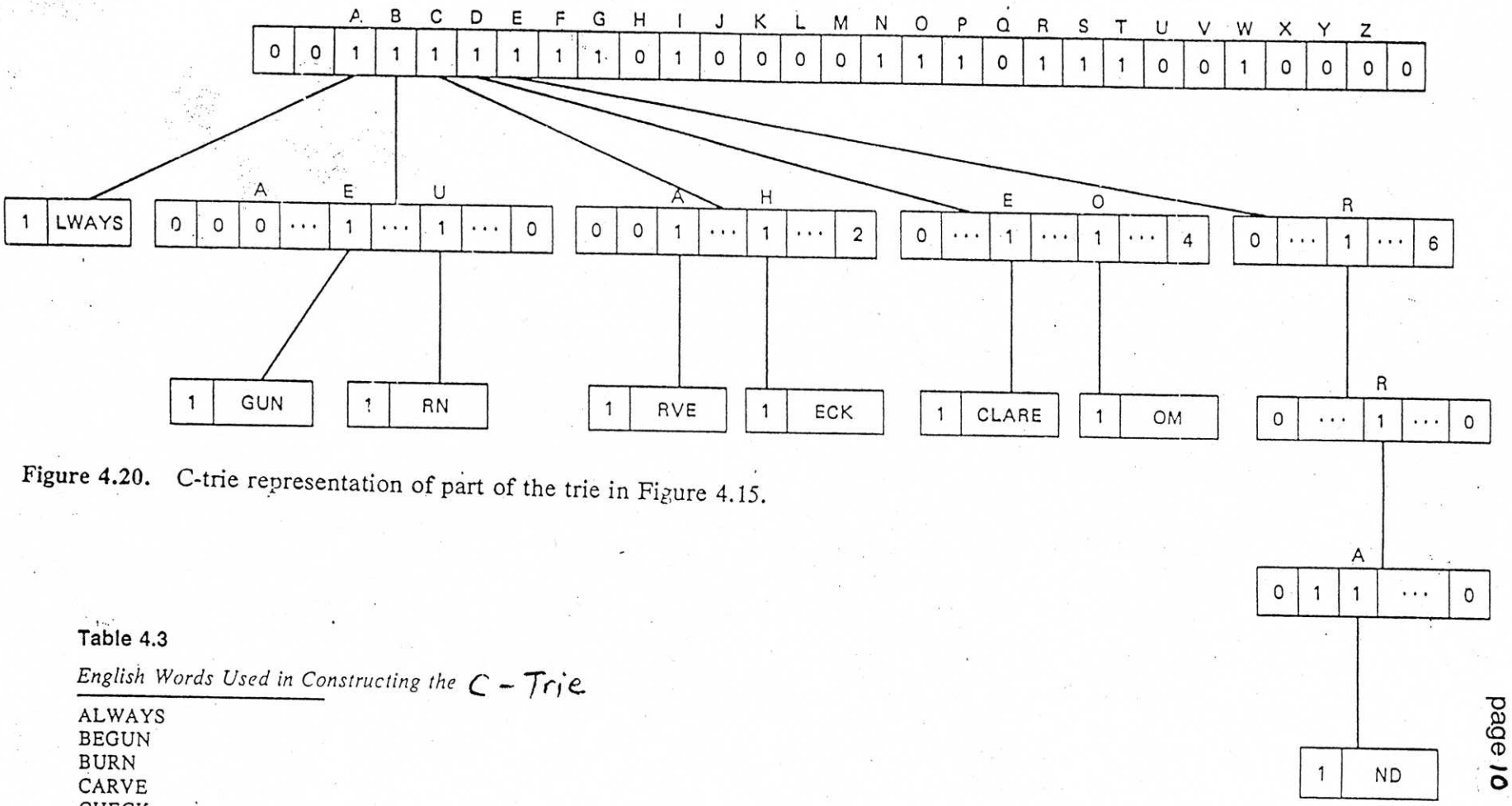

Figure 4.20. C-trie representation of part of the trie in Figure 4.15.

Table 4.3

English Words Used in Constructing the C - Trie

ALWAYS
BEGUN
BURN
CARVE
CHECK
DECLARE
DOOM
ERR
ERRAND